





HD28
.M414
v. 3779
95

Dewey



Image-Information Systems for Traffic Management

Ichiro Masaki
Amar Gupta

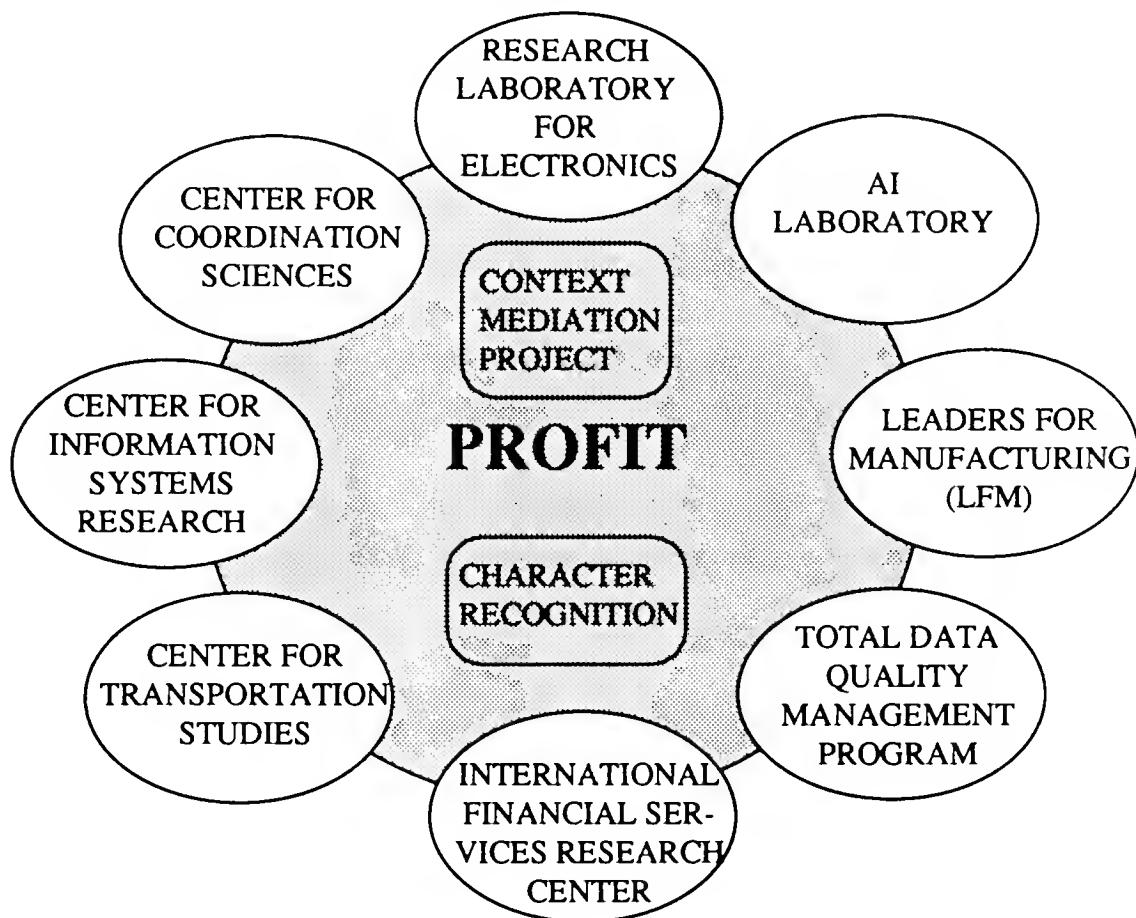
WP #3779 August 1994
PROFIT #94-18

Productivity From Information Technology
"PROFIT" Research Initiative
Sloan School of Management
Massachusetts Institute of Technology
Cambridge, MA 02139 USA
(617)253-8584
Fax: (617)258-7579

Copyright Massachusetts Institute of Technology 1994. The research described herein has been supported (in whole or in part) by the Productivity From Information Technology (PROFIT) Research Initiative at MIT. This copy is for the exclusive use of PROFIT sponsor firms.

Productivity From Information Technology (PROFIT)

The Productivity From Information Technology (PROFIT) Initiative was established on October 23, 1992 by MIT President Charles Vest and Provost Mark Wrighton "to study the use of information technology in both the private and public sectors and to enhance productivity in areas ranging from finance to transportation, and from manufacturing to telecommunications." At the time of its inception, PROFIT took over the Composite Information Systems Laboratory and Handwritten Character Recognition Laboratory. These two laboratories are now involved in research related to context mediation and imaging respectively.



In addition, PROFIT has undertaken joint efforts with a number of research centers, laboratories, and programs at MIT, and the results of these efforts are documented in Discussion Papers published by PROFIT and/or the collaborating MIT entity.

Correspondence can be addressed to:

The "PROFIT" Initiative
Room E53-310, MIT
50 Memorial Drive
Cambridge, MA 02142-1247
Tel: (617) 253-8584
Fax: (617) 258-7579
E-Mail: profit@mit.edu

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

MAY 26 1995

LIBRARIES

Image-Information Systems for Traffic Management

Ichiro Masaki and Amar Gupta

Sloan School of Management

Masachusetts Institute of Technology

Abstract

This paper describes some examples of image-information systems which are relevant to traffic management. After reviewing related work in the fields of traffic management, intelligent vehicles, stereo vision, and ASIC-based approaches, the paper focuses on a stereo vision system for intelligent cruise control. The system measures the distance to the vehicle in front using trinocular triangulation. An application specific processor architecture was developed to offer low mass-production cost, real-time operation, low power consumption, and small physical size. The system was installed in the trunk of a car and evaluated successfully on highways.

1. Introduction

Advanced traffic management systems and intelligent vehicle systems are expected to serve as an important component of the social infrastructure for the next generation. The recent rapid surge in interest in this field is motivated by two factors: technical and social. The technical reason is that the underlying technologies such as signal processing, communication, computers, and sensors have finally reached the level at which the intelligent-vehicle-related devices can be produced at affordable prices. For example, when General Motors and RCA jointly demonstrated an intelligent cruise control system during the fifties, no one saw any possibility of converting it into a real product. The system used a microwave radar to measure the three dimensional positional relationship between the automated vehicle and the vehicle in front. The steering and the

A revised version of this paper will be submitted to SPIE Photonics East '94.

An earlier version was presented at the IEEE IECON '93 conference in November, 1993.

speed were automatically controlled so that the automated vehicle followed the vehicle in front. The system was too big and the predicted production cost made commercial exploitation infeasible. With significant technology progresses made in the related areas, the intelligent cruise control system, for example, is expected to be on the market in the near future.

The social reasons for the rapid increase of interest in intelligent vehicles include the fact that conventional approaches for improving highway traffic systems are reaching their limits and we need a new approach. In the past, we expanded the highway systems physically to accommodate ever-increasing traffic congestion. It is, however, getting increasingly difficult to build additional lanes to existing highways or to make new highways because of environmental concerns, costs, and other issues. Intelligent vehicle-highway systems are expected to offer a new approach for developing safe, highly-efficient, and environmentally-friendly traffic systems.

Intelligent vehicle-highway systems are based on various perception capabilities, and can be categorized into two groups: on-vehicle systems and on-road systems. Examples of on-vehicle systems include ones for intelligent cruise control, collision warning, collision avoidance, warning for lane changing, obstacle detection for backing up, and side collision prediction for exploring side air bags. The on-road systems include ones for traffic flow monitoring for controlling traffic flow, vehicle identification for automated toll gates, and vehicle monitoring for identifying speeding vehicles automatically.

2. ASIC-based Vision for Intelligent Vehicles

The sensing approaches being studied both for the on-vehicle and on-road systems include vision, microwave (millimeterwave), acoustic, and laser radar. The advantages of the vision based approaches are as follows:

- (1) Only vision systems offer the potential for measuring lane boundaries without changing existing roads. Lane sensing is useful for a number of applications. With intelligent cruise control systems, for example, the lane sensing capability makes it possible to measure the distance to the car in front in the same lane, not in the next lane, even on curves.
- (2) Since no sensing method is perfect at this time, it is important that drivers can predict and understand when the system might fail. The vision systems offer characteristics similar to human visual perception, making it easier to predict when they might fail.
- (3) Vision systems are passive and do not emit anything, making it unnecessary to consider health, regulation, and interference issues.

Potential drawbacks of the vision systems are:

- (1) The vision systems do not work when human drivers cannot see; whether this is a drawback or a merit is controversial because it would be dangerous if people could drive at high speeds in the dense fog based on excessive reliance on the microwave radar.
- (2) The history of research in vision-based approaches is shorter than that with microwaves.

Vision systems can be categorized into two groups: ASIC-based and microprocessor-based systems. With ASIC-based schemes, the major visual processing tasks are performed by ASIC (Application Specific Integrated Circuit) chips. The microprocessor-based schemes, in contrast, use general-purpose microprocessors as the major components. Many systems combine these two schemes at various degrees. The most significant merit of the ASIC-based methods is that their product designs are highly efficient [1-3] offering higher processing speeds, lower power consumptions, and smaller silicon area sizes compared to microprocessor-based methods. The ASIC architectures do not offer unnecessary flexibility for specific applications and therefore much higher processing speeds can be obtained with smaller silicon areas. The processing speed increased by a factor of 2×10^2 to 1×10^3 compared to the conventional microprocessor-based approaches [4]. On the other side, ASIC-based approaches involve high development costs and deliver less flexibility. Development costs are high because ASIC schemes need to develop dedicated IC chips, instead of using off-the-shelf microprocessor chips.

ASIC approaches, therefore, are appropriate for the systems to be mass-produced where the size of the market compensates for the high development costs and less flexibility, and also in situations requiring high processing speeds and low power consumptions. These conditions apply to intelligent vehicle applications an annual automobile production rate of 40-50 million units world-wide provides large-market opportunities to products related to intelligent vehicles. Also the vision-based vehicle guidance applications require low production costs, high processing speeds, and low power consumption rates.

3. Related Work

3.1 Vision Systems for Intelligent Vehicles

A number of examples of vision systems for intelligent vehicles are described in [5] and [6]. All of them are still in the research stage. Only a few types of similar sensors (e.g.,

millimeter and laser systems) are commercially available for warning purposes. This section describes some examples of vehicle applications of vision systems.

One popular application is road following or lane following. The vision system measures the curvature of the road or the lane in front and controls the steering angle and the vehicle speed to follow the road/lane automatically. A research group at Universitat der Bundeswehr Munchen uses six microprocessor boards in parallel for real-time detection of the road edges and each of these processors is dedicated to a specific region of the image frame. A number of vision systems for road following have been developed at Carnegie Mellon University, including one that uses neural network technology. Lane markings themselves, instead of lane edges, are used as explicit objects by researchers at Bristol University. The lane markings are extracted based on knowledge of their size, shape, and gray-level intensity characteristics. Road models are then used to verify candidates of lane markings. Assuming that the road surface is on a single straight plane with a single radius; these road models are simpler than others such as the one developed at University of Maryland.

At Yamanashi University, a lit road segment is merged with a shaded road segment using the normalized red and green intensities; that is, the percentages of the red and green light vectors in the total light. In this process, the lit and shadowed portions of the road can be merged into the same road region although these portions are different in terms of the intensities. Extremely dark shadows, however, cannot be processed with this method because the dynamic range of television cameras is not large enough to provide reliable normalized color information for really dark areas. Other approaches for road/lane detection include texture-based method developed at Laboratoire Heudiasyc and hybrid approach involving neural network methodology and texture-based segmentation from Universitat Politecnica de Catalunya.

Experiments at Matsushita Corporation indicate that the reliability of the visual recognition of the lanes depends heavily on the weather conditions. For example, success rates are as high as 97% during day-time and 98% during night-time with fine or cloudy weather, but only 26% at sunrise and sunset. A research group at Mazda Corporation sees a necessity for further research in vision systems and knowledge-based reasoning systems for automatic lane following on real highways; they report as much as 5% failure in recognizing lane marks on real highways because of a variety of external disturbances. Knowledge-based reasoning capability is important in deciding actions based on the external information acquired by vision systems. Other institutions which have published papers on road/lane following include General Motors, Nissan, Honda, National Institute of Standards and Technology in the U.S. Department of Commerce.

Traffic sign recognition is another application field. A system from Daimler-Benz involves three steps for recognizing traffic signs. In the first step, color segmentation is performed using neural networks. The second step involves generation of hypotheses on the image region

containing traffic signs and the kind of the signs based on prior knowledge on the traffic signs and involves outdoor scenes; this whole knowledge is stored in a frame-based network. The third and final step evaluation of the hypotheses and outputting the result.

Peugeot has developed a road sign recognition system in which road signs are classified into three categories depending on the contour shapes: octagonal, triangular, and circular for stops, danger warnings, and less important information respectively. Both the octagonal stop signs and triangular danger signs include red color to facilitate detection. A monochrome video camera, installed near the rear view mirror, contains an optical filter for reducing red light in order to increase the contrast between the red regions and the white borders in the signs. Closed contours are extracted from the binary edge image and are represented in the Freeman code format. For classification, a neural network approach was chosen over an expert system or a structured programming method because the neural net approach required a shorter development time and a shorter processing time. Experiments were done at medium speeds, 40 to 60 km/h, and most signs were recognized.

Adaptive cruise control is an extension of a conventional cruise control system in which the engine throttle is controlled for maintaining vehicle speed constant. The adaptive cruise control system adjusts the speed depending on the speed of the vehicle in front and other factors. The function to follow a vehicle in front is an important part of the adaptive cruise concept. A car-following system developed at Ruhr-Universität Bochum takes a symmetric object in an image as the back view of the vehicle in front. The system performs both tracking and identification of the object. The edge image of the object is correlated with deformable two dimensional models using an elastic net technique.

Obstacle detection is useful for adaptive cruise and collision avoidance. A feature of Renault's system is that it includes two cameras: a usual video camera and a near-infrared camera. The system, therefore, offers high sensitivity to red lights including tail lights of the car in front. Daihatsu Corp. has combined two-camera stereo and optical flow methods for obstacle detection. While the "obstacle" usually means a slow going vehicle in front, researchers at Universität der Bundeswehr München have developed a system to detect vehicles approaching from behind. At University of Massachusetts, a group of vision algorithms have been developed for intelligent vehicle applications including lane change/merge warning, automatic lane changing system, side collision warning, automatic collision avoidance, vision enhancement, intersection hazard warning, and lateral control (steering control).

3.2 Stereo Vision

Three-dimensional vision systems can be classified into two categories: indirect and direct systems [7]. The indirect systems use single images and calculate distances based on the focus information or other information. The direct systems include time-of-flight and stereo methods. The stereo approaches are based on triangulation, and can be classified into three categories: active stereo using laser, passive stereo involving multiple images, and optical flow including a time factor. Motion stereo is an example of well-known stereo schemes [16-17] and a 3D data of objects is calculated from a sequence of monocular images. Some of the motion stereo systems belong to the passive stereo and use discrete features such as lines and corners for 3D calculations, while other stereo methods use optical flow [18]. [19] is an example of systems which include both static stereo and motion stereo features. A potential problem with the motion stereo for intelligent vehicles is that one cannot assume that the motion of the camera is known. Another factor in classifying three-dimensional vision systems is whether geometric models of objects are known or not [20-21].

A significant problem with binocular stereo vision involves finding what part in the right image corresponds to what part in the left image. Even an axial layout in which two cameras share the same optical axis cannot solve this correspondence problem without assuming some constraints [22]. One approach to solve this problem is to assume some constraints like Marr-Poggio-Grimson algorithm, and a second approach is to use symbolic representation for matching [23]. [29] classifies binocular stereo vision into two categories: area-based and feature-based systems. Area-based stereo systems offer the advantage of directly generating a dense disparity map but are sensitive to noise and breakdown where there is a lack of texture or where depth discontinuities occur. Feature-based systems, in contrast, are less sensitive to noise and highly accurate in the depth measurement but provide only sparse depth map and handle the smoothness assumption with difficulty. Trinocular stereo vision increases the geometric constraints and reduces the influence of heuristic constraints for stereo-matching. [8] contains a list of references on trinocular stereo vision, including [9].

Many stereo research projects focus on the algorithm aspects with less emphasis on real-time processing. The processing time depends on various factors such as complexity of images and models of computers, and the computation-intensive nature of stereo algorithms is exemplified by numbers such as 174 s and 14.5 s [11], 10 min. [23], and 1 hr and 5 hrs [29]. Special architectures are needed to shorten the processing times and some examples of such architectures are described in the following section.

3.3 Architecture for Vision Processors

Vision systems require high speed processing [14-15]. The architectures for low level visual processing can be classified into three types: parallel binary array processors, pipelined processors, and special function units. The parallel binary array processor consists of a large number of bit-serial processing elements and near neighbor interconnection of the processing elements. In many applications, each processing element corresponds to one pixel and the array processor works in a SIMD (Single-Instruction Multi-Data) scheme. The bit-serial architecture allows flexible data formats and makes the system very efficient with respect to memory and processing resource utilization. Many image processing algorithms require combination of data within local areas of each pixel; the near neighbor interconnection scheme enables these algorithms to be implemented efficiently. Examples of this type of processors include CLIP4, MPP, and associative processors.

Pipeline processors take image data in a raster scan format from the television camera or the image storage memory into the first stage of the pipeline. In the initial setup mode, the host computer specifies the function of each stage through the instruction bus and loads the whole pipeline with the image data. An N -by- N -pixel image data, therefore, requires N -by- N clock cycles, plus initial setup time, to complete the process. This type of processor does not require a high speed controller because the controller does not have to change the instructions for processing elements after the initial setup. The cytocomputer and FLIP system are examples of this type of parallel processors for visual processing.

With special function units scheme, each special function unit represents a direct hardware implementation of a visual image algorithm, or in some cases to a set of related functions. Special function units may contain some local memory and program control. These units are usually connected through a high speed bus or interconnection networks. The inner product computer (IPC) and TOSPICS are examples of this type of visual processors. This paper focuses on this category.

Some examples of ASIC chips for intelligent vehicle applications are described in [5]. These examples include local pattern processor (LPP) chip that performs 2D convolutions with a programmable kernel at a TV-rate (6 MHz) for edge detection and other applications, and Hough parameter estimator (HPE) chip designed for real-time Hough transformation.

An emerging technology in the field of ASIC chips for vision is the analog vision chip scheme. An overview of developments in this area at the Massachusetts Institute of Technology is presented in [27], this includes seven different analog chips for image filtering and edge detection, moment extraction to determine object position and orientation, image smoothing and

segmentation, depth determination from stereo image pairs, accurate depth determination jointly from imperfect depth and slope data, and camera motion determination, plus additional chips to test novel circuit designs and processing methods. Potential merits of analog vision chips, as compared to digital chips, include small silicon areas, high speeds, and low mass-production costs.

4. Stereo Vision for Intelligent Cruise Control

4.1 Stereo System Architecture

A desired vision system can be designed in either scale-down or scale-up mode. In the scale-down approach, a vision system that produces the desired result is developed without considering the cost and the processing time in the first development stage, and the cost and the processing time are reduced in the following stages. In the scale-up approach, in contrast, one develops a system that delivers the best performance within the prescribed processing time and system cost, and the performance is improved in the following stages. The later approach was selected to develop the first prototype which works in real-time with a reasonable system size. Our strategy for the first step was to develop a simple, small, real-time stereo vision system for intelligent cruise control and to try it on highways to find real problems.

Figure 1 shows the system block-diagram. The total process consists of the following four steps:

- (1) Image acquisition: Three cameras take images simultaneously.
- (2) Feature extraction: Features are extracted from three intensity images. The features are vertical edge segments.
- (3) Stereo matching: The feature images are shifted each other for disparity measurements. Two binocular stereo pairs, with three cameras, eliminate most false correspondences.
- (4) Post filtering: Post filtering eliminates distance information generated by non-vehicle objects such as lane markings.

These four steps are described in more details below.

In the image acquisition step, three CCD (Charge Coupled Device) television cameras are installed in the front of the car. The right and left cameras are each separated from the central camera by 30 cm. In the second step, positive and negative vertical edges are calculated from the three camera images. Pixels at which the intensity levels increase or decrease significantly from the left to the right of those pixels are defined as positive and negative edges, respectively. We

calculate only vertical edges and ignore horizontal edges because the three cameras are located on a single horizontal line. The binary edges are processed by a segment filter. The filter eliminates edges unless they are part of five-pixel-long vertical edge segments. If four or five pixels in the five-pixel-long segment are edge pixels, it is considered a valid edge segment. The vertical segments are used as features for stereo matching.

In the third step, stereo matching is carried out between the right and center images and between the center and left images in parallel. This dual-matching approach eliminates a significant portion of the false correspondences. The right and left images are shifted one column by one column to the left and right, respectively, a matched trio situation exists when three corresponding pixels, each of which is in the right, center, and left image respectively, are all positive or all negative edges. Since the correlation peak of binary edge correlation is very sharp, the edge width of the center image was extended to three pixels while the right and left images have one-pixel-wide edges. Pixels which have matches at multiple disparity values are calibrated at the nearest distance values based on safety considerations. The output of the third step is a distance map which indicates a distance value at every matched pixel. In the final stage, a histogram is calculated from the distance map image. The horizontal and vertical axes of the histogram are the distance value and the number of pixels which belongs to each distance value. The histogram is then self-convoluted with a window for some distance, for example +/- one shift distance, so that the new histogram represents the number of pixels which belong to each distance range that overlaps with each other. The system recognizes the peak as the nearest object if the peak value (i.e., the number of pixels) is larger than the threshold value, as shown in Figure 2. Through this process, the edge pixels which represent lane markings, for example, are filtered out because they do not make any significant peaks.

4.2 ASIC-based Approach for Feature Detection

The features used in our first version are vertical edge segments. An algorithm for feature detection is described below. First, a 3x3-pixel Sobel-Ratio operator for vertical edges calculates spatial intensity gradient values as follows.

3x3-pixel window:		NW	N	NE
		W	C	E
		SW	S	SE

If $(NE+2E+SE) \geq (NW+2W+SW)$,

$$\text{Gradient Value} = (NE+2E+SE)/(NW+2W+SW)$$

Otherwise,

$$\text{Gradient Value} = (NW+2W+SW)/(NE+2E+SE)$$

One merit of the Sobel-Ratio operator over the conventional Sobel operator is that the gradient values depend only on the reflection rates of objects and not on the ambient light level. With the Sobel operator, in contrast, the gradient values are defined as the absolute values of the differences of pairs of column sums: $(NE+2E+SE)$ and $(NW+2W+SW)$. The Sobel values depend on products of the reflection ratios and the ambient light level, and therefore the threshold level for binarization must be adjusted depending on the ambient conditions. The Sobel-Ratio operator, however, does not work well in dark regions where both column sums are small and the signal-to-noise ratios are not good. Figure 3 describes our hardware implementation for the feature detection. In the spatial gradient unit, the 3×3 -pixel window is made with two line buffers. Two sets of adder circuits calculate two column sums in parallel. The upper 7 bits of each column sum are used for spatial gradient calculation. The 14-bit data specifies the address of a $16K \times 7$ -bit PROM (Programmable Read Only Memory). All the spatial gradient calculations, including the conditional jump for choosing Sobel-Ratio or Sobel, the calculation of Sobel-Ratio or Sobel, and the scaling of the calculated result, are implemented in this PROM.

The 7-bit output of the PROM goes to the second processing unit. The second unit, or a thinning/thresholding unit, generates single-bit-wide binary edges from the spatial gradient images. The pixels which satisfy the following two conditions become binary edges:

- (1) The gradient values are larger than the threshold value.
- (2) The gradient values are local maxima.

This processing step was implemented using comparators.

The third processing unit for feature calculation consists of a vertical segment filtering unit. This unit eliminates all edge pixels which are not parts of the vertical edge segments; as a part of a vertical edge segment is assumed to exist if four or more pixels in a five-pixel-long vertical window are edge pixels. In addition to North- and South-connections, the definition of the vertical connections includes all the 45-degree-connections such as North-East and South-West-connections. This process eliminates isolated edge pixels and cleans up the binary edge images. The unit consists of two blocks: a window block and a logic tree block. The window block

captures binary edges within a certain region using line buffers, and sends them to the logic tree block. The logic tree block consists of logical-AND gates and logical-OR gates to track connectivity of edge pixels on a vertical basis.

The algorithm described above, differs from conventional sophisticated algorithms [10, 12-13, 25-26, 28]; the former focuses on efficient use of silicon, whereas the latter focuses on speed. For example, the edge segment filtering is carried out in the binary edge domain in our implementation, whereas, from a pure algorithm point of view, a gray-level spatial gradient image is preferred for extracting edge segments and offers results better than without algorithm. Questions asked during the development of this system were: "How much additional silicon areas do we need for how much performance improvements?" and; "How important is performance improvement of each sub-system from a total system point of view?" In the case of edge segment filtering, for example, it was easy to justify the cost of extending a single bit window block to a 7-bit one. The silicon area size of the logic tree block, however, would be increased by a factor which is significantly larger than seven if we would convert the edge representation from binary to 7-bit. A significant silicon area increase would be expected because the conventional algorithms require a conditional-branch-type logic instead of a straight-forward logic which was implemented for the first version.

The above discussion exemplifies the nature of the ASIC-based approach in which the algorithm and hardware design are verified at various design stages from a specific application point of view. The algorithm should be compared with alternative ones with various criteria including total system performance and required silicon areas. The extracted binary edge segments are transferred to the stereo matching unit described in the following section.

4.3 ASIC-based Approach for Stereo Matching

The stereo matching unit receives six images from the feature extraction units. The six images consist of three positive and three negative edge segment images. Each of the right, center, and left video cameras generates a pair of positive and negative edge segment images. The stereo matching unit, therefore, consists of two identical sub-systems: one each for the positive and the negative edge segments. These two sub-systems work in parallel and their outputs are combined to generate a single distance map. The stereo matching custom processor shifts the right and left images one column by one column to the left and the right, respectively. Cross-correlation among the center, shifted-right, and shifted-left images are calculated at each shift value. Edge pixels in the center image are tagged with the shift value if the corresponding pixels both in the shifted-right and shifted-left images are edge pixels.

Figure 4 describes the hardware implementation for stereo matching. The positive and negative edge segment matching operations are performed by two sets of 8-parallel-processors. The system therefore includes 16 of three-input-logical-AND-gates: eight gates each for the positive and negative edge segments. Corresponding binary edge segment data from the right, center, and left images go to these three input ports of each gate. If these three data values are all high, it is considered as a matched case and the corresponding disparity value is recorded at the corresponding pixel location in the distance map. The distance map pixels which have multiple matches store the largest disparity, or the shortest distance, for a safety reason. The above procedure is shown below in greater details:

Input RP(n,m), CP(n,m), and LP(n,m) to Gate-1.

If the output of Gate-1 is high, write "disparity-0" at D(n,m).

Input RP(n+1,m), CP(n+1,m), and LP(n+1,m) to Gate-2.

If the output of Gate-2 is high, write "disparity-0" at D(n+1,m).

.....

.....

Input RP(n+7,m), CP(n+7,m), and LP(n+7,m) to Gate-8.

If the output of Gate-7 is high, write "disparity-0" at D(n+7,m).

Input RP(n,m-1), CP(n,m), and LP(n,m+1) to Gate-1.

If the output of Gate-1 is high, and D(n,m) is non-zero, write high at M(n,m).

If the output of Gate-1 is high, write "disparity-1" at D(n,m).

.....

.....

Input RP(n+7,m-1), CP(n+7,m), and LP(n+7,m+1) to Gate-8.

If the output of Gate-7 is high, and D(n+7,m) is non-zero, write high at M(n+7,m).

If the output of Gate-7 is high, write "disparity-1" at D(n+7,m).

where,

RP(n,m), CP(n,m), and LP(n,m): data at row-n, column-m of positive edge line segment image from right, center, and left cameras, respectively.

D(m,n): data at row-n, column-m of the distance map

M(n,m): data at row-n, column-m of the multiple-match-flag map. "high" for multiple-match and 'low" for single-match or no-match

4.4 Evaluation

The first version of the ASIC-based processing units for feature extraction and stereo matching was implemented into custom boards (one extended VMS board for each unit) using off-the-shelf CMOS logic chips such as logical-AND-chips and counter-chips to evaluate the architecture. The speeds of this ASIC-based architecture are compared to microprocessor-based approaches in this section.

The number of operations for feature detection is calculated as follows:

Number of binary edge detection: 256×256 [pixels/image] $\times 3$ [images for right, center, and left] $\times 10$ [times/sec] = 2 M [pixel-operations/sec]

Each pixel operation includes the following 16 arithmetic/logic operations:

4 additions for calculating two column sums

2 comparisons and 2 conditional branches for choosing Sobel-Ratio or Sobel operator

1 comparison and 1 division for Sobel-Ratio operation

3 comparisons and 3 conditional jumps for thinning

Each binary edge requires the following 486 arithmetic/logic operations:

324 additions, 81 comparisons, and 81 conditional jumps for line-segment-filtering

If 5% of all the pixels are binary edge pixels, the total number of arithmetic/logic operations required for the feature detection is calculated as follows:

$$2M \times (16 + 486 \times 5/100) \times 1.5 = 121 M \text{ arithmetic/logic operations}$$

In the above estimation, a factor of 150% was utilized to include extra operations such as read-data, write-data, calculate-addresses, and other minor operations.

Suppose that the machine cycle of a microprocessor is 30 MHz and that every arithmetic/logic operation requires two machine cycles in average, the speed of the ASIC system is about 8 times faster than a microprocessor-based implementation. The size of ASIC feature detection board is similar to that of an off-the-shelf single board microprocessor. If the board size

of our system would be reduced by a factor of 100 by replacing low-density small-scale off-the-shelf logic chips with application-specific VLSI (Very Large Scale Integration) chips, the speed/size ratio of the ASIC-based system would be better by a factor of 800 as compared to a microprocessor-based system.

The following portion of this section discusses the evaluation of the ASIC-based stereo matching architecture. The application specifications for intelligent cruise control requires the following conditions for the stereo matching unit:

Feature image size: 256 x 256 pixels
Disparity range: 0 - 96 pixels
Processing speed: depth map every 100 msec

The calculation area is 256x256 pixels for disparity-0 (zero) decreased by two for each disparity increment. With disparity-96, the calculation area is (256 - 2x96) x 256 pixels. Two identical series of operations are required: one each for positive and for negative edge segment images. The total number of matching operations is calculated as follows:

$$((256 \times 256 + (256-2x96) \times 256)/2) [\text{operations/disparity}] \times 97 [\text{disparities}] \times 2 [\text{/stereo_match}] \times 10 [\text{stereo_matches/sec}] = 79 \text{ M} [\text{matching operations/sec}]$$

If each matching operation requires four machine cycles, on average, for logical-AND, address-calculation, data-write, data-read, and other operations, then a speed of 320 MHz is required; this is 11 times of the typical 30 MHz speed. Since the board size of our first version is twice as large as that of the typical microprocessor board, the ASIC-based scheme would be about 550-times superior in terms of the speed/size ratio, the board size of our system was reduced by a factor of 100 by replacing the off-the-shelf small-scale arithmetic/logic chips with ASIC chips.

5 . Closing Remarks

In the context of advanced traffic systems including intelligent vehicles, vision-based systems offer significant advantages over other approaches. Accordingly, the first version of an intelligent cruise control system was developed using stereo vision techniques and implemented with low-density small-scale off-the-shelf chips. The results of on-highway experiments indicated acceptable performance. The second version now under development at MIT, replaces the original

digital scheme with a hybrid analog/digital scheme with the objective of lowering ultimate production cost.

Acknowledgments

The authors are indebted to G. G. Dodd, R. B. Tilove and other individuals at General Motors for developing the first version. Special thanks to B. K. P. Horn, H-S. Lee, C. Sodini, J. White, J. L. Wyatt, and other members of the MIT analog vision chip group for the interesting discussions.

References

- (1) Seitz, C.L., "Concurrent VLSI Architectures", *IEEE Transactions on Computers*, vol. c-33, no.12, December 1984
- (2) Schmitt, L.A., WILSON, S.S., "The AIS-5000 Parallel Processor", *IEEE Transactions on Pattern and Machine Intelligence*, vol. 10, no.3, May 1988
- (3) Vick, C.R., Kartashev, S.P., Kartashev, S.I., "Adaptable Architectures for Supersystems", COMPUTER MAGAZINE, PP. 17-35, NOVEMBER 1980
- (4) Masaki, I., "Industrial Vision Systems Based on Application-Specific IC Chips", *IEICE Transactions on Electronics*, vol. E74, no. 6, June 1991
- (5) Masaki, I., Vision-based Vehicle Guidance, Springer-Verlag, 1991
- (6) Masaki, I. - Editor, Proceedings of IEEE Intelligent Symposium '92, IEEE, July, 1992
- (7) Nitzan, D., "Three-Dimensional Vision Structure for Robot Applications", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no.3, May 1988
- (8) Ayache, N., Lustman, F., "Trinocular Stereo Vision for Robotics", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 1, Jan. 1991
- (9) Yachida, M., "3-D Data Acquisition by Multiple Views", *Robotics Research: Third Int. Symp.*, Faugeras, O.D. and Giralt, G., Eds. Cambridge, MA: MIT Presss, 1986, pp. 11-18
- (10) Wilson, R., Bhalerao, A.H., "Kernel Designs for Efficient Multiresolution Edge Detection and Orientation Estimation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 3, March 1992
- (11) Horaud, R., Skordas, T., "Stereo Correspondence Through Feature Grouping and Maximal Cliques", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 11, Nov. 1989

- (12) Canny, J., "A Computational Approach to Edge Detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. Pami-8, no.6, Nov. 1986
- (13) Davis, L.S., "A Survey of Edge Detection Techniques", *Computer Graphics and Image Processing* 4, pp. 248-270, 1975
- (14) Masaki, I., "SEAMSIHT: A Parallel/Pipelined Vision System for Seam Tracking", *Proc. IEEE 7th International Conference on Pattern Recognition*, pp. 424-427, 1984
- (15) Masaki, I., "Parallel/Pipelined Processor Dedicated to Visual Recognition", *Proc. 1985 IEEE International Conference on Robotics and Automation*, pp. 100-107, 1985
- (16) Leung, M.K., Huang, T.S., "An Integrated Approach to 3-D Motion Analysis and Object Recognition", *Machine Intelligence*, vol. 13, no. 10, Oct. 1991
- (17) Tirumalai, A.P., Schunck, B.G., Jain, R.C., "Dynamic Stereo with Self-Calibration", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.14, no. 12, Dec. 1992
- (18) Young, G.-S. J., Chellappa, R., "3-D Motion Estimation Using a Sequence of Noisy Stereo Images: Models, Estimation, and Uniqueness Results", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no.8, Aug. 1990
- (19) Zhang, Z., Faugeras, O.D., "Estimation of Displacements from Two 3-D Frames Obtained from Stereo", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 12, Dec. 1992
- (20) Linnainmaa, S., Harwood, D., Davis, L.S., "Pose Determination of a Three-Dimensional Object Using Triangle Pairs", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, no. 5, Sept. 1988
- (21) Maitre, H., Luo, W., "Using Models to Improve Stereo Reconstruction", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, Feb. 1992
- (22) Alvertos, N., Brzakovic, D., Gonzalez, R.C., "Camera Geometries for Image Matching in 3-D Machine Vision", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 9, Sept. 1989
- (23) Grimson, W.E.L., "Computational Experiments with a Feature Based Stereo Algorithm", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. Pami-7, no. 1, Jan. 1985
- (25) Fleck, M.M. "Some Defects in Finite-Difference Edge Finders", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 3, March 1992
- (26) Tan, H.L., Gelfand, S.B., Delp, E.J., "A Cost Minimization Approach to Edge Detection Using Simulated Annealing", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 1, Jan. 1991

- (27) Wyatt, J. L., Jr., Keast, C., Seidel, M., Standley, D., Horn, B., Knight, T., Sodini, C., Lee, H-S., and Poggio, T., "Analog VLSI Systems for Image Acquisition and Fast Early Vision Processing", *International Journal of Computer Vision*, p.217-230, 1992
- (28) Jeong, H., Kim, C.I., "Adaptive Determination of Filter Scales for Edge Detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no.5, May 1992
- (29) Cavanagh, P., "Reconstructing the Third Dimension: Interactions between Color, Texture, Motion, Binocular Disparity, and Shape", *Computer Vision, Graphics, and Image Processing* 37, pp. 171-195, 1987

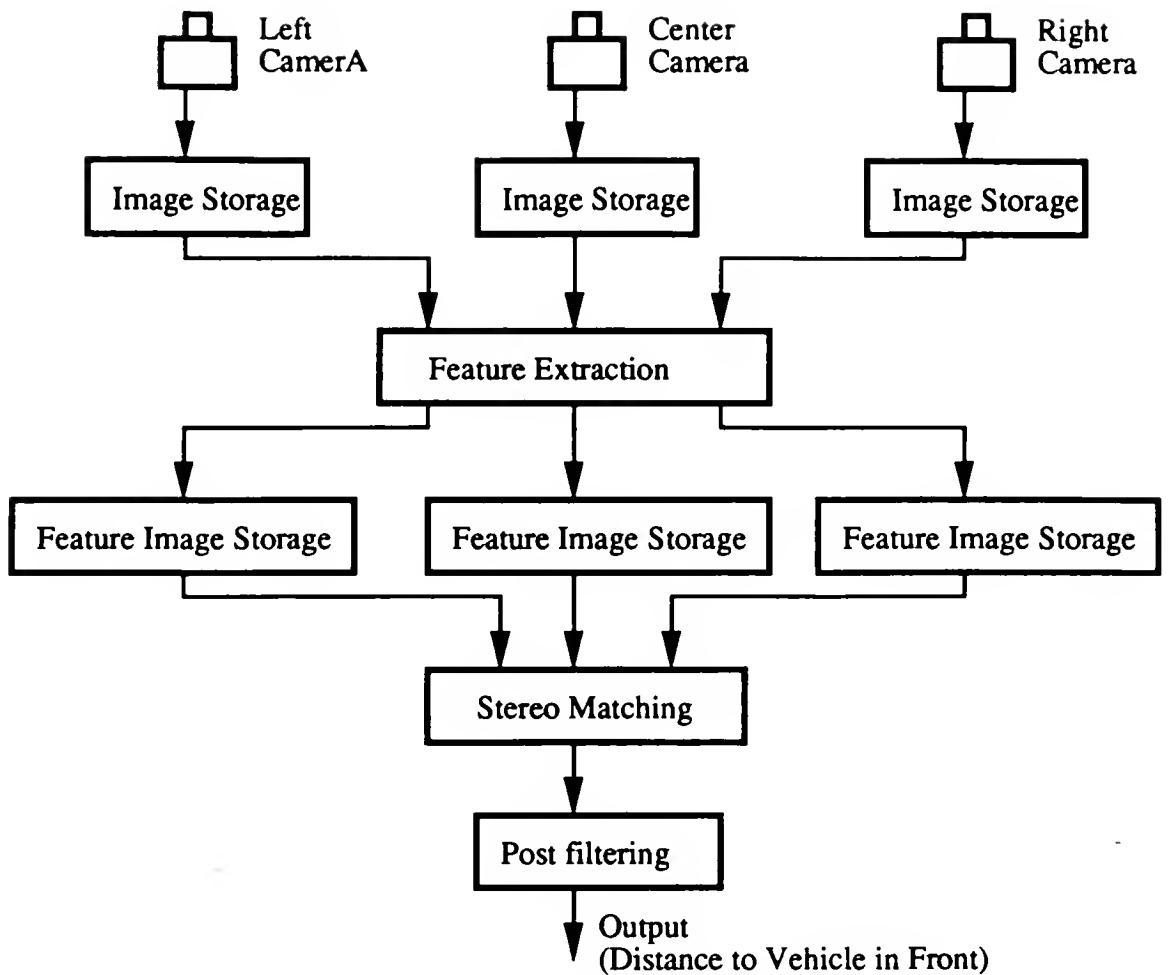


Figure 1. Block-Diagram of Stereo System

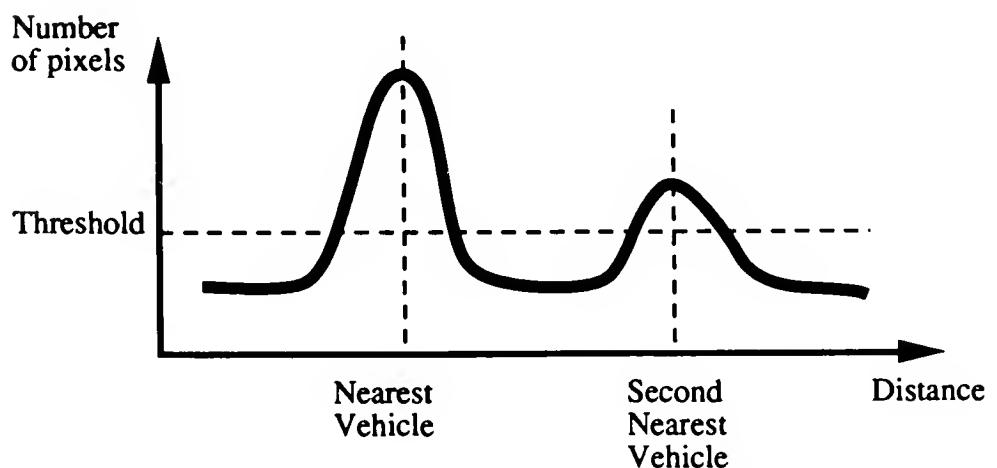


Figure 2. Post Filtering Operation

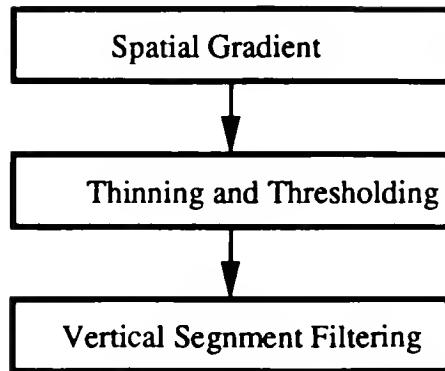


Figure 3. Feature Detection Hardware

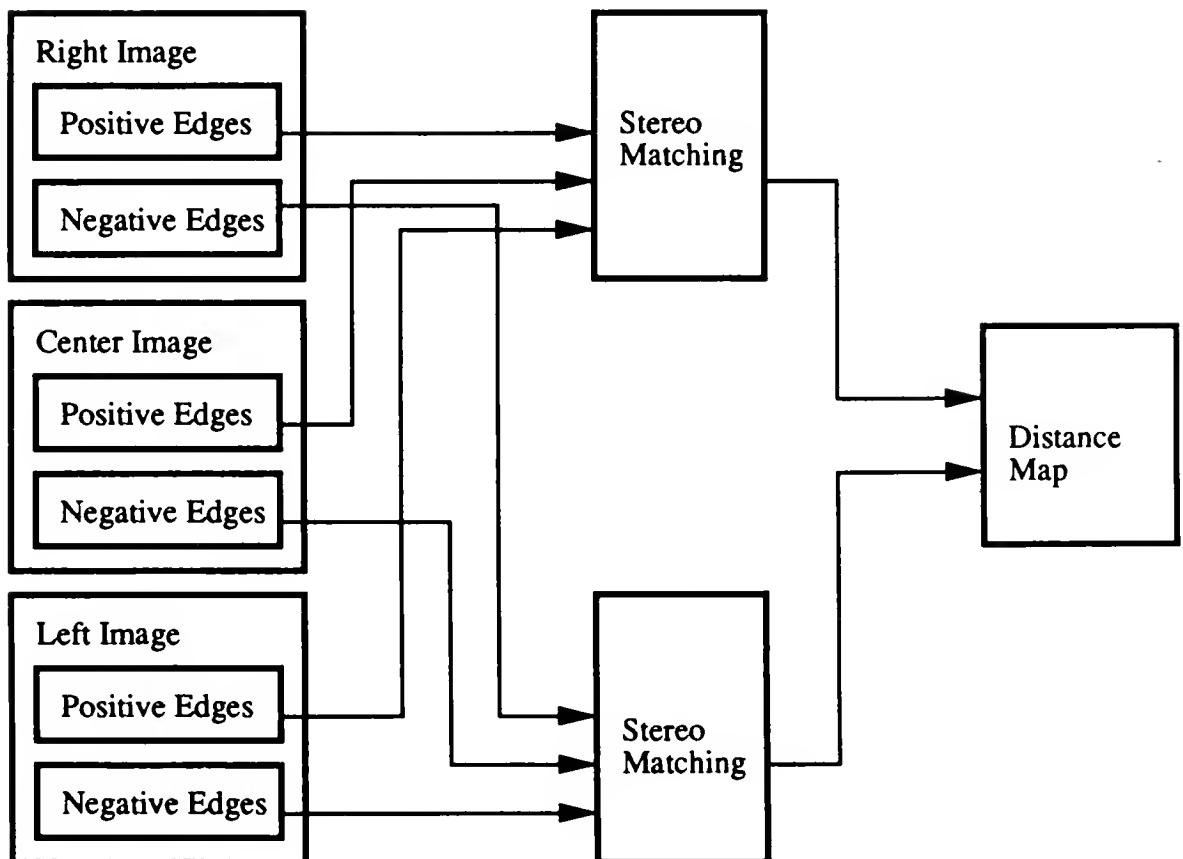


Figure 4. Stereo Matching Hardware

MIT LIBRARIES
3 9080 00932 7302

Date Due

		Lib-26-67

